Cairo University
Faculty of Graduate Studies for Statistical Research

Enhancing Road Damage Detection by Optimizing Faster R-CNN with ResNet-50 Backbone

A Project Presented For fulfillment.
For a Master project in Software Engineering

Submitted by:

Kahlawy S. Hussien

Supervised by:

Dr. Ashraf A. Shahin

2024

Enhancing Road Damage Detection by Optimizing Faster R-CNN with ResNet-50 Backbone

Kahlawy Hussein *(Author)*
Department of Computer Since
Faculty of Graduate Studies
for Statistical Research
Cairo University, Egypt
12422023437012@pg.cu.edu.eg

***Abstract*—Road damage poses significant challenges to road safety and transportation efficiency. Traditional road damage detection methods are time-consuming, labor-intensive, and prone to human error. Recent advancements in deep learning have enabled automated and accurate road damage detection from visual data. This research proposes a novel deep learning framework for enhancing road damage detection by optimizing the Faster R-CNN architecture with a ResNet-50 backbone. A carefully curated dataset of 500 images, split into training and validation sets, is utilized. The images are preprocessed using data augmentation techniques to improve model robustness. The Faster R-CNN model is initialized with pre-trained weights and fine-tuned on the road damage dataset. The training process involves optimizing the model using the SGD optimizer and adjusting the learning rate based on validation loss. The model's performance is evaluated using the Mean Average Precision (mAP) metric on the validation set. The experimental results demonstrate the effectiveness of the proposed framework, with the model achieving promising results in accurately detecting and localizing road damage. The developed model has the potential to support timely road maintenance and infrastructure assessment by enabling efficient identification of road damage. This research contributes to the advancement of automated road damage detection techniques and highlights the significance of deep learning in addressing this critical challenge.**

Keywords—Road damage detection, deep learning, Faster R-CNN, ResNet-50, data augmentation, Mean Average Precision (mAP)

Introduction

Road damage, Road infrastructure plays a vital role in the economic and social development of countries worldwide. However, the deterioration of road surfaces, including the formation of potholes and other types of damage, poses significant challenges to road safety, ride comfort, and transportation efficiency [1]. Traditional road damage detection methods rely on manual inspection or specialized equipment, which can be time-consuming, labor-intensive, and prone to human error [2, 3]. These limitations highlight the need for automated, accurate, and efficient road damage detection systems to support timely maintenance and repair efforts. In recent years, the rapid advancements in computer vision and deep learning have opened up new opportunities for automating road damage detection using various data sources, such as smartphone images [1, 4], unmanned aerial vehicle (UAV) imagery [6, 9, 10, 13], and crowdsourced data [7, 12]. Deep learning algorithms, particularly convolutional neural networks (CNNs), have demonstrated remarkable performance in detecting and classifying road damage from visual data [1, 3, 5, 14]. These algorithms can learn hierarchical features from raw images and accurately localize and categorize diverse types of road damage, enabling more efficient and effective road maintenance strategies. Researchers have explored various deep learning architectures and techniques to enhance the performance and generalizability of road damage detection models. These include the adaptation of popular object detection frameworks such as YOLO [2, 4, 8, 9], Faster R-CNN [9], and SSD [1], as well as the incorporation of attention mechanisms [5, 15], transfer learning [4], data augmentation [2, 8], and lightweight architectures [15]. Moreover, the development of large-scale road damage datasets, such as the Road Damage Dataset (RDD) [1, 7, 8, 12], the China Road Damage Detection Dataset (CNRDD) [5], and the Pavement Image Dataset (PID) [14], has facilitated the training and evaluation of deep learning models for this task. Despite the substantial progress made in automated road damage detection, several challenges remain to be addressed. These include the need for more efficient and lightweight models suitable for real-time deployment on resource-constrained devices, the development of robust models that can handle diverse road conditions and damage types across different geographical regions, and the integration of these models into practical road maintenance and asset management systems. This research paper aims to contribute to the advancement of automated road damage detection by proposing a novel deep-learning framework that leverages state-of-the-art techniques to achieve high accuracy and real-time performance. The proposed framework combines the Faster R-CNN architecture with a ResNet-50 backbone, data augmentation techniques, and a carefully curated dataset to develop a robust and efficient pothole detection model. The model's performance is evaluated on multiple benchmark datasets, demonstrating its effectiveness in detecting and classifying various types of road damage. The remainder of this paper is organized as follows: Section 2 provides an overview of the related works in the field of road damage detection using deep learning. Section 3 describes the methodology employed in this study, including the dataset preparation, model architecture, training process, and evaluation metrics. Section 4 presents the experimental results and analysis, showcasing the model's performance and comparing it with state-of-the-art methods. Finally, Section 5 concludes the paper and discusses potential future research directions.

Background

Road damage, particularly potholes and surface defects, poses significant challenges to road safety, ride comfort, and transportation efficiency [1]. Traditional road damage detection methods rely on manual inspection or specialized equipment, which can be time-consuming, labor-intensive, and prone to human error [2, 3]. In recent years, advancements in computer vision and deep learning have paved the way for automated road damage detection using various data sources, including smartphone images [1,4], unmanned aerial vehicle (UAV) imagery [6,9,10,13], and crowdsourced data [7,12]. Deep learning algorithms, particularly convolutional neural networks (CNNs), have shown remarkable performance in detecting and classifying road damage from visual data [1, 3, 5, 14]. Various architectures, such as YOLO [2, 4, 8, 9], Faster R-CNN [9], and SSD [1], have been adapted and optimized for road damage detection tasks. These models can learn hierarchical features from raw images and accurately localize and categorize diverse types of road damage [1, 3, 5, 14]. To further enhance the performance and generalizability of deep learning models, researchers have explored techniques such as transfer learning [4], data augmentation [2, 8], attention mechanisms [5, 15], and lightweight architectures [15]. Large-scale road damage datasets have been introduced to facilitate the development and evaluation of deep-learning models. Notable examples include the Road Damage Dataset (RDD) [1, 7, 8, 12], the China Road Damage Detection Dataset (CNRDD) [5], and the Pavement Image Dataset (PID) [14]. In this research, we focus on the detection and classification of potholes and road damage using the "potholes_and_road_damage_with_annotations" dataset. The dataset provides annotations in the Pascal VOC format, which includes five types of potholes and road damage. By leveraging this dataset, we aim to develop a deep learning framework that can accurately detect and classify various types of potholes and road damage.The proposed framework utilizes the Faster R-CNN architecture with a ResNet-50 backbone, which has shown promising results in object detection tasks [9]. We explore techniques such as data augmentation and transfer learning to enhance the model's performance and robustness. The model is trained and evaluated on the "potholes_and_road_damage_with_annotations" dataset, and its effectiveness is assessed using standard evaluation metrics such as mean Average Precision (mAP). This research aims to contribute to the advancement of automated road damage detection by proposing a deep learning framework tailored to the specific task of pothole and road damage detection. The practical implications of this work include improved road maintenance and infrastructure assessment, enabling timely identification and repair of potholes to ensure road safety and minimize vehicle damage.

Related Works

Road damage detection is a critical task for ensuring road safety and efficient maintenance. Traditional methods rely on manual inspection or specialized equipment, which can be time-consuming, expensive, and prone to human error. In recent years, deep learning techniques have emerged as a promising solution for automated road damage detection using smartphone imagery. This literature review presents the state-of-the-art deep learning approaches for road damage detection and classification.

Maeda et al. (2018)[1], addressed the problem of road damage detection using deep learning techniques. Previous studies have focused on detecting the presence or absence of damage, but not on classifying the specific types of damage. Additionally, there is no standard benchmark dataset for road damage detection, making it difficult to compare different approaches. The methodology of this research involves three main contributions. First, the authors create a large-scale road damage dataset consisting of 9,053 images with 15,435 instances of damage. The images were captured using a smartphone mounted on a car dashboard in collaboration with 7 municipalities in Japan. The dataset covers a diverse range of weather and illumination conditions. Each damage instance is annotated with a bounding box and classified into one of 8 categories. Second, they train state-of-the-art object detection models (SSD with Inception V2 and SSD with MobileNet) on this dataset and provide benchmark results. Third, they demonstrate high accuracy in classifying damage into 8 categories. The results show that the SSD with MobileNet model achieves the best performance, with recall and precision greater than 71% and 77%, and running at 1.5 seconds per image on a smartphone [1].

Wan et al. (2022)[2], developed an automated solution for detecting and classifying diverse types of road damage from smartphone images. The problem they address is efficiently monitoring urban road conditions by recognizing specific damage types, which is crucial for planning effective maintenance efforts. Their methodology involves employing a deep learning approach based on the YOLO (You Only Look Once) object detection algorithm. They treat each road damage type as a distinct object and train the YOLO detector on a dataset of annotated images categorized into eight damage classes defined by the Japan Road Association. To manage class imbalance, they augment the training data by synthesizing additional images for underrepresented classes. The trained model takes an image as input and outputs the detected damage types along with their locations. The authors evaluate their approach on a test set of 1,813 images and report an F1-score up to 0.62 for their best-performing model, trained on the augmented dataset. They find that increasing the non-maximum suppression value, which allows more overlapping predicted boxes, slightly improves the F1-score. Their solution and code are openly available for further research and development. In summary, the authors propose a deep learning-based object detection approach for automated road damage classification from smartphone images, addressing an important problem in urban infrastructure maintenance. The authors employ state-of-the-art techniques, including data augmentation and parameter tuning, to achieve promising results on a challenging dataset [2].

Mraz et al. (2020)[3], focused on the problem of detecting and localizing road damage using deep neural networks. Road damage, such as cracks and potholes, poses safety risks and incurs high maintenance costs. Traditional road damage detection methods rely on manual inspection or specialized equipment, which can be time-consuming and expensive. The paper aims to develop an automated and efficient solution for road damage detection using deep learning techniques. The methodology employed in this research paper involves the development of a localized road damage detection model using deep neural networks. The authors utilize a dataset of road damage images captured by a vehicle-mounted camera. The dataset is preprocessed and augmented to increase its size and diversity. The proposed model architecture consists of a convolutional neural network (CNN) for feature extraction and a fully connected layer for damage localization. CNN is trained using the augmented dataset, and the model's performance is evaluated using metrics such as precision, recall, and F1-score. The model's ability to detect and localize road damage has been compared with other state-of-the-art methods. The authors present the results of the developed localized road damage detection model. The model achieves high accuracy in detecting and localizing diverse types of road damage, such as cracks and potholes. The authors provide quantitative evaluations of the model's performance, including precision, recall, and F1- score. The results demonstrate that the proposed deep neural network approach outperforms traditional methods and achieves state-of-the-art performance in road damage detection. The paper also discusses the potential applications of the developed model, such as automated road maintenance and asset management systems. The authors highlight the scalability and efficiency of the proposed approach, making it suitable for large-scale road damage detection tasks. [3].

Arya et al. (2020)[4], presented a research problem addressed in this paper is the need for a reliable and efficient system for detecting road damage in real-time. The authors argue that this system is crucial for the success of autonomous driving technology. To address this issue, the researchers proposed a transfer learning-based detection method, which involves using pre-trained networks to accurately classify road damages. The system utilizes two pre-trained models for feature extraction and classification respectively, and they are fine-tuned for road damage detection. The experimental results showed that the proposed method performed well in detecting 7 types of road damage with an average precision rate of 93.54%. The method also achieved reliable results when assessed on different datasets, proving its robustness. Overall, the authors demonstrated the feasibility and potential of transfer learning-based methods for road damage detection. In summary, the authors presented a transfer learning-based method for road damage detection, illustrating its effectiveness and reliability through experiments. The findings suggest that this system can integrated into autonomous vehicles, which can improve the safety and efficiency of driving.[4]

Zhang et al. (2022)[5], addressed the challenge of automated road damage detection using deep learning. introduce the China Road Damage Detection Dataset (CNRDD), which contains 4,319 images with 21,552 damage annotations across 8 categories and 3 severity levels. The CNRDD dataset is more comprehensive and challenging compared to existing datasets, such as RDD2020, due to its larger number of categories, higher damage density per image, and inclusion of severity levels. The authors propose a novel deep-learning framework for road damage detection that incorporates attention mechanisms. The framework consists of an attention fusion module, which uses edge detection to guide the network's attention to salient damage regions, and a salient feature learning module with normalization, which suppresses the weights of non-salient features to improve damage discrimination. A two-stage coarse-to-fine training strategy is employed to first learn features broadly and then fine-tune the specific damage categories. Evaluations of the CNRDD and RDD2020 datasets demonstrate the effectiveness of the proposed method, which outperforms state-of-the-art damage detection models. The proposed method achieves F1-scores of 34.73% and 51.72% on the CNRDD and RDD2020 datasets, respectively. Analysis reveals that cracks tend to be the easiest damage type to detect. The authors discuss future directions to advance the field, including standardizing datasets, optimizing similarity measures between damage features, and conditional transfer learning across domains. Continuously updating the CNRDD will also provide more training data to further improve automated road damage detection [5].

Silva et al. (2023)[6], focused on the problem of automated road damage detection using Unmanned Aerial Vehicle (UAV) images and deep learning techniques. The manual collection of road damage data is labor-intensive and potentially unsafe for humans. Therefore, the authors propose using UAVs and Artificial Intelligence (AI) technologies to significantly improve the efficiency and accuracy of road damage detection. The proposed approach utilizes three deep learning algorithms, YOLOv4, YOLOv5, and YOLOv7, for object detection and localization in UAV images. The authors trained and tested these algorithms using a combination of the RDD2022 dataset from China and a Spanish road dataset. They also implemented a YOLOv5 model with a Transformer Prediction Head to address the issue of large variations in object scales. The dataset was augmented and preprocessed to increase its size and diversity, and the models were evaluated using metrics such as precision, recall, mean average precision (mAP), and inference time. The experimental results demonstrate that the proposed approach is efficient and achieves high accuracy in detecting road damage. The YOLOv5 version achieved a mAP0.5 of 59.9%, while the YOLOv7 version achieved a mAP0.5 of 73.20%. The YOLOv5 model with a Transformer Prediction Head achieved a mAP0.5 of 65.70%. The authors also conducted a visual analysis of the results, which showed that the models could accurately identify and locate road damage structures in UAV images. These results demonstrate the potential of using UAVs and deep learning for automated road damage detection and pave the way for future research in this field [6].

Saha et al. (2024)[7], addressed the problem of developing robust road damage detection models that can identify and classify different types of road damage across multiple regions or countries. Training such models typically require extensive, diverse, and well-labeled data, which can be difficult due to data privacy concerns, large data transfers, storage needs, and computational resources. Federated learning (FL) is explored as a solution to train models collaboratively without the need for data sharing, by only exchanging model parameters between clients and the server. The methodology consists of five steps. First, road damage data sets are selected and prepared for single-country (Japan) and multi-country (Japan, India, and the United States) scenarios, with both independent and identically distributed (IID) and non-IID data distributions. Second, the Flower framework is chosen for deploying FL. Third, object detection using YOLOv5l is performed with FL on Flower. Fourth, centralized and FL models are trained for both single-country and multi-country scenarios. Finally, the performance of the models is evaluated using mean average precision (mAP) and compared, considering factors such as data distribution, data volume, server rounds, and local epochs. The results show that FL models have 21%–25% lower mAP compared to centralized models, with data distribution significantly influencing the performance of FL models. The FL model trained on non-IID multi-country data outperformed local country models by 1.33%–163% when evaluated on global test data, demonstrating the potential for developing more robust and generalizable road damage detection models. The study also found that countries without their own road damage data sets can leverage FL models trained on multiple countries to perform road damage detection without accessing data from those countries. Factors such as the number of clients, server rounds, and local epochs were found to affect the performance and training time of FL models [7].

Mizanur & Mustakim (2023)[8], presented the problem of road damage detection using deep learning techniques. Road damage poses significant risks to public safety and can lead to increased transportation costs, traffic disruptions, and economic losses. Traditional road inspection methods are time-consuming, expensive, and prone to human error. Therefore, there is a need for an automated, accurate, and efficient method to detect and classify road damage using computer vision and deep learning algorithms. The methodology employed in this research paper involves using the YOLOv7 deep learning model for road damage detection. The authors conducted three experiments to compare the performance of YOLOv5 and YOLOv7 models. In Experiment 1, they trained the YOLOv7 model using transfer learning and various image augmentation techniques, such as rotation, scaling, size adjustment, and left-right flipping. Experiment 2 introduced the Gaussian Blur method for image augmentation, while Experiment 3 utilized the YOLOv5 algorithm with a refined dataset. The dataset used in this study is the Road Damage Dataset (RDD2022), which contains 47,420 images from six countries, with more than 55,000 instances of road damage annotated into eight classes. The results of the three experiments demonstrate that the YOLOv7 model, trained using the methodology in Experiment 1, exhibits the highest accuracy and viability compared to the other models. The proposed model achieved an average accuracy of 79.75% across all test image categories, with an F1-score of 75%. This is a 1.5% improvement over the state-of-the-art model. The authors also developed a user-friendly computer application with a graphical user interface (GUI) to facilitate the convenient implementation of the proposed model by end-users. The adoption of this technology in practical applications is justified by its exceptional performance in providing significant insights into road conditions and enabling timely maintenance and repair interventions [8].

Chen et al. (2024)[9], presented the problem of developing a systematic solution for automatic crack detection and pavement distress evaluation using Unmanned Aerial Vehicles (UAV) inspection systems. Traditional manual inspection methods are time-consuming, labor-intensive, and prone to human error, while vehicle-mounted inspection systems are expensive and limited in scope. The authors propose a comprehensive framework that integrates UAV data acquisition, deep learning-based crack identification, and road damage assessment to enhance the practicality and widespread application of UAV inspection systems for pavement distress detection. The methodology employed in this study involves several key components. First, a flight control strategy is established to ensure high-quality pavement imagery acquisition by the UAV, utilizing a flight suitability parameter model. The acquired UAV imagery is pre-processed through frame extraction, image dividing, and data enhancement, followed by labeling according to five major categories of cracks. Four mainstream deep learning target detection algorithms (Faster-RCNN, YOLOv5, YOLOv7, and YOLOv8) are then trained and evaluated using the labeled dataset. The best-performing model is selected based on precision, recall, F1-score, and mean accuracy precision (mAP) metrics. Finally, the preferred model is employed to identify road crack targets in UAV imagery, and quantitative assessments are conducted to evaluate pavement distress. The experimental results demonstrate the effectiveness of the proposed framework for automatic crack detection and pavement distress evaluation using UAV inspection systems. The Faster-RCNN model achieves the highest accuracy and effectiveness in recognizing fine cracks, with a precision of 75.6%, recall of 76.4%, F1-score of 75.3%, and mAP of 79.3%. The YOLO series models, while slightly less accurate, exhibit significant advantages in terms of training speed and low video memory requirements. Among the YOLO models, YOLOv5s and YOLOv8s show comparable recognition accuracy, while YOLOv7-tiny performs the worst. The study also highlights the importance of UAV data acquisition quality, with the proposed self-made crack dataset outperforming existing datasets in terms of crack recognition accuracy and algorithmic efficiency. Furthermore, the authors present quantitative measurement methods for road cracks and demonstrate their application in evaluating pavement distress, providing factual evidence for maintenance decisions made by road authorities [9].

Merkle et al. (2022)[10], addressed the problem of automatically assessing road surface conditions from aerial imagery using deep learning techniques. Conventionally, road condition assessment is carried out using terrestrial sensors like laser scanners or line scan cameras, which provide accurate data but are time-consuming and costly. The authors investigate the feasibility of using aerial imagery, which offers a more efficient and large-scale alternative, to extract two key parameters for evaluating the condition of asphalted roads: cracks and working seams. The methodology employed in this study involves adapting two fully convolutional neural network architectures, namely Dense-U-Net and Skip Fuse-Dense-U-Net, for the task of extracting thin road surface features from aerial images. The authors experiment with various training setups, including the use of road masks as additional input, targeted training on patches overlapping with road areas, and the application of a smooth mean squared error loss function to handle the thinness of the target objects. A new dataset consisting of 132 labeled aerial images with a ground sampling distance (GSD) between 2-10 cm is generated to train, validate, and test the models. The results demonstrate that the proposed deep learning models are capable of effectively extracting cracks and working seams from aerial imagery. The best-performing Skip Fuse-Dense-U-Net architecture achieves an intersection over union (IoU) of 46.82% for cracks and 31.85% for working seams, with even higher scores of 63.98% and 45.58%, respectively, when using a 3-pixel tolerant IoU. These findings indicate the potential of using state-of-the-art segmentation models for fast and large-scale assessment of road surface conditions. The authors suggest future work to test the method in other areas, extend it to include additional relevant objects, and investigate the influence of factors such as GSD, acquisition time, weather, and seasonal differences on the quality of the results [10].

Rodriguez-Lozano et al. (2023)[11], addressed the problem of high dimensionality in pavement crack classification algorithms, which leads to complex models, reduced classification accuracy, and increased computation time. Automatic crack classification plays a crucial role in road maintenance, but using many features for classification is inefficient and challenging for embedded systems with limited computational resources. The methodology proposed in this paper is a new data dimensionality reduction method called DDR4CC (Data Dimensionality Reduction for Crack Classification Algorithms). DDR4CC reduces the required information about cracks to only four interpretable features: $\{\max V,\ dV,\ \max H,\ dH\}$. These features are not influenced by the spatial location of the defect in the images and can be extracted regardless of the original image resolution. DDR4CC is compared with eight other dimensionality reduction methods, and the reduced set of features is analyzed using five different classification algorithms across five datasets generated by combining several public datasets. The experimental results demonstrate that DDR4CC enhances the performance of the classification algorithms, providing nearly perfect classifiers with minimal computation time. The proposed method outperforms other dimensionality reduction techniques in terms of classification accuracy and execution time. DDR4CC achieves the highest F1-score and the lowest execution time when combined with various classification algorithms, making it the most efficient method for crack classification on low-computational capacity systems. The authors conclude that DDR4CC's interpretable features and fast computation enable real-time detection and classification on low-consumption devices, which could be particularly useful for unmanned aerial vehicles (UAVs) in road maintenance applications [11].

Arya et al. (2022)[12], focused on the Crowdsensing-based Road Damage Detection Challenge (CRDDC'2022), which aims to address the problem of automatically detecting and classifying road damages using images collected from six countries: India, Japan, the Czech Republic, Norway, the United States, and China. The challenge builds upon the previous Road Damage Detection Challenge (RDDC'2018) and Global Road Damage Detection Challenge (GRDDC'2020), which focused on road damage detection in Japan and three countries, respectively. The main objective of CRDDC'2022 is to develop more robust and generalizable models by expanding the scope to six countries and allowing participants to contribute their datasets. CRDDC'2022 employed a four-phase methodology to achieve its objectives. In Phase 1, participants were invited to submit their road damage datasets, which were then shortlisted for inclusion in the challenge during Phase 2. The main task of Phase 3 required participants to propose road damage detection models using the RDD2022 dataset, which consisted of 47,420 images from the six target countries. The dataset was divided into training and testing sets, with annotations provided only for the training set. Participants trained their models and submitted predictions for the test set, which were evaluated using an online server. Five leaderboards were maintained to assess model performance for each country and overall. Finally, Phase 4 required the submission of detailed reports and source code for the proposed solutions. recommendations for future challenges, such as allocating more time for experiments, providing computational resources to participants, allowing the use of external data sources, and considering additional evaluation metrics. CRDDC'2022 successfully advanced the state-of-the-art in road damage detection by providing a diverse, multi-national dataset and fostering the development of robust, generalizable models through a phased data science competition [12].

Pan et al. (2018)[13], addressed the problem of automated detection of asphalt pavement potholes and cracks using unmanned aerial vehicle (UAV) multispectral imagery and machine learning algorithms. The methodology employed in this study involves several key steps. First, multispectral imagery (12 bands from 450-1000 nm) was acquired using a UAV platform, providing rich spectral information compared to standard RGB images. Next, multi-resolution segmentation was applied to generate image objects, and features including spectral means and standard deviations, geometric properties, and texture measures were extracted for each object. Three machine learning classifiers - Support Vector Machine (SVM), Artificial Neural Network (ANN), and Random Forest (RF) - were then trained and compared using various feature combinations. The impact of spatial resolution on detection accuracy was also investigated. The results demonstrate that the Random Forest classifier with 18 trees, utilizing a combination of spectral, geometric, and texture features, achieved the highest overall accuracy of 98.3% in detecting potholes, cracks, and non-distressed pavement. Geometric and texture features were found to be more important than spectral features alone. The study also highlighted the importance of spatial resolution, with cracks being missed when the pixel size exceeded 3 cm. In conclusion, UAV multispectral imaging combined with machine learning provides an efficient and accurate tool for monitoring asphalt pavement conditions, supporting timely maintenance decision-making. This research contributes to the field by demonstrating the effectiveness of advanced imaging and classification techniques for automated pavement distress assessment [13].

Majidifard et al. (2020)[14], addressed the problem of automating pavement distress detection and developing a comprehensive pavement condition index using deep learning techniques. Traditional manual pavement surveys are time-consuming, expensive, and prone to human subjectivity. Previous automated methods have limitations in accurately classifying multiple distress types and quantifying their severity. There is a need for a low-cost, robust system that can detect, classify, and quantify pavement distress to support pavement management decisions. The methodology involves several key steps. First, a large dataset of 7,237 pavement images was compiled from Google Street View and manually annotated with bounding boxes for nine distress types. This Pavement Image Dataset (PID) was used to train deep-learning models. A pre-trained YOLO model was used for distress detection and classification. A U-Net model was employed for pixel-level distress segmentation to quantify density and severity. The outputs of the YOLO and U-Net models were then combined into a comprehensive hybrid model. Finally, various machine learning techniques, including Genetic Expression Programming (GEP), linear regression, and a PCI-based weighting method, were used to develop pavement condition index prediction models based on the distress type and density information from the deep learning models. The results demonstrate the effectiveness of the deep learning-based pavement evaluation system. The YOLO model achieved high accuracy (F1-score of 0.84) in classifying the nine distress types. The U-Net model accurately segmented distresses, including in challenging shadowed conditions. The hybrid model leveraging both YOLO and U-Net outputs enabled a comprehensive assessment considering both distress type and density. The pavement condition index models showed a strong correlation with ground truth PASER values, with R^2 values up to 0.94 on test sections. The GEP model identified block cracking as the most influential distress. Overall, this study presents a novel, automated, cost-effective, and robust pavement evaluation methodology that can support timely and reliable maintenance decisions. With larger training datasets, the models can be further enhanced for even broader deployment [14].

Chen et al. (2024)[15], addressed the problem of road damage detection, which plays a crucial role in ensuring road safety and improving traffic flow. Traditional manual inspection methods are time-consuming, labor-intensive, and prone to human error. With advancements in computer vision and machine learning, automatic road damage detection has become a more efficient and accurate approach. However, most existing deep learning-based methods suffer from slow inference speed and large model size, thus failing to satisfy real-time detection and deployment on smaller embedded traffic devices. The methodology proposed in this paper is a lightweight attention ghost-you-only look once (LAG-YOLO) network for efficient road damage detection. LAG-YOLO optimizes the network structure of YOLO, making it more suitable for real-time processing and lightweight deployment while ensuring high accuracy. The authors introduced a novel attention ghost module to reduce the model parameters and improve the model performance. The ghost module ensures the model has a small memory footprint and low computational complexity, while the simple attention module (SimAM) enhances the model's ability to focus on important information. The proposed method is evaluated on two datasets: RDD2020 and Hualu. The experimental results demonstrate the effectiveness of LAG-YOLO for road damage detection. On the RDD2020 dataset, LAG-YOLO achieves an impressive mean average precision (mAP) score of 52.35% with only 4.19 million parameters, outperforming several state-of-the-art methods such as CenterNet, Faster R-CNN, and RetinaNet. Similarly, on the Hualu dataset, LAG-YOLO obtains (mAP) 45.80%, surpassing the original YOLOv5 while maintaining a lightweight architecture. The proposed method achieves real-time performance with 50 frames per second (FPS) and requires significantly fewer parameters and floating-point operations per second (FLOPs) compared to other methods. These results indicate that LAG-YOLO is a promising solution for efficient road damage detection, suitable for deployment in practical scenarios with limited computational resources [15].

Table 1: Comparing Road Damage Detection algorithms with the results.

No.	Research Paper	Algorithms	Results
1	Chen et al. (2024) [15]	Lightweight attention ghost-you only look once (LAG-YOLO) network	RDD2020 dataset: mAP score of 52.35% with 4.19 million parameters. Hualu dataset: mAP of 45.80%. Real-time performance with 50 FPS and fewer parameters and FLOPs compared to other methods.
2	Chen et al. (2024) [9]	Faster-RCNN, YOLOv5, YOLOv7, and YOLOv8	Faster-RCNN achieves the highest accuracy with precision of 75.6%, recall of 76.4%, F1-score of 75.3%, and mAP of 79.3%. YOLOv5s and YOLOv8s show comparable recognition accuracy, while YOLOv7-tiny performs the worst.
3	Saha et al. (2024) [7]	Federated learning with YOLOv5l for object detection	FL models have 21%–25% lower mAP compared to centralized models. FL model trained on non-IID multi-country data outperformed local country models by 1.33%–163% on global test data.
4	Rodriguez-Lozano et al. (2023) [11]	DDR4CC (Data Dimensionality Reduction for Crack Classification Algorithms)	DDR4CC enhances the classification algorithm's performance, providing nearly perfect classifiers with minimal computation time. Outperforms other dimensionality reduction techniques in terms of accuracy and execution time.
5	Silva et al. (2023) [6]	YOLOv4, YOLOv5, and YOLOv7 for object detection and localization in UAV images	YOLOv5: mAP of 59.9%, YOLOv7: mAP of 73.20%, YOLOv5 with Transformer Prediction Head: mAP of 65.70%.
6	Mizanur & Mustakim (2023) [8]	YOLOv7 with transfer learning and image augmentation techniques	The average accuracy of 79.75% across all test image categories, with an F1-score of 75%. 1.5% improvement over the state-of-the-art model.
7	Arya et al. (2022) [12]	Crowdsensing-based Road Damage Detection Challenge (CRDDC'2022)	Advances state-of-the-art in road damage detection by providing a diverse, multi-national dataset and fostering the development of robust, generalizable models through a phased data science competition.
8	Merkle et al. (2022) [10]	Dense-U-Net and Skip Fuse-Dense-U-Net for extracting thin road surface features from aerial images	Skip Fuse-Dense-U-Net: IoU of 46.82% for cracks and 31.85% for working seams, with higher scores of 63.98% and 45.58% using a 3-pixel tolerant IoU.
9	Wan et al. (2022) [2]	YOLO object detection with data augmentation	F1-score up to 0.67 for the best-performing model.
10	Zhang et al. (2022) [5]	Deep learning framework with attention fusion module and salient feature learning module	F1-scores of 34.73% and 51.72% on CNRDD and RDD2020 datasets, respectively, outperforming state-of-the-art damage detection models.
11	Arya et al. (2020) [4]	Transfer learning with pre-trained networks for feature extraction and classification	The average precision rate of 93.54% in detecting 7 types of road damage. Robust results on different datasets.

12	Majidifard et al. (2020) [14]	YOLO for distress detection and classification, U-Net for pixel-level distress segmentation, and machine learning techniques for pavement condition index prediction	YOLO: F1-score of 0.84 in classifying nine distress types. U-Net: accurate distress segmentation. Pavement condition index models showed a strong correlation with ground truth PASER values, with R^2 values up to 0.94 on test sections.
13	Mraz et al. (2020) [3]	CNN for feature extraction and fully connected layer for damage localization	High accuracy in detecting and localizing various types of road damage, outperforming traditional methods.
14	Pan et al. (2018) [13]	Support Vector Machine (SVM), Artificial Neural Network (ANN), and Random Forest (RF) with UAV multispectral imagery	Random Forest with 18 trees, utilizing spectral, geometric, and texture features, achieved 98.3% overall accuracy in detecting potholes, cracks, and non-distressed pavement.
15	Maeda et al. (2018) [1]	SSD with Inception V2 and SSD with MobileNet	SSD with MobileNet: recall and precision greater than 71% and 77%, running at 1.5 seconds per image on a smartphone.

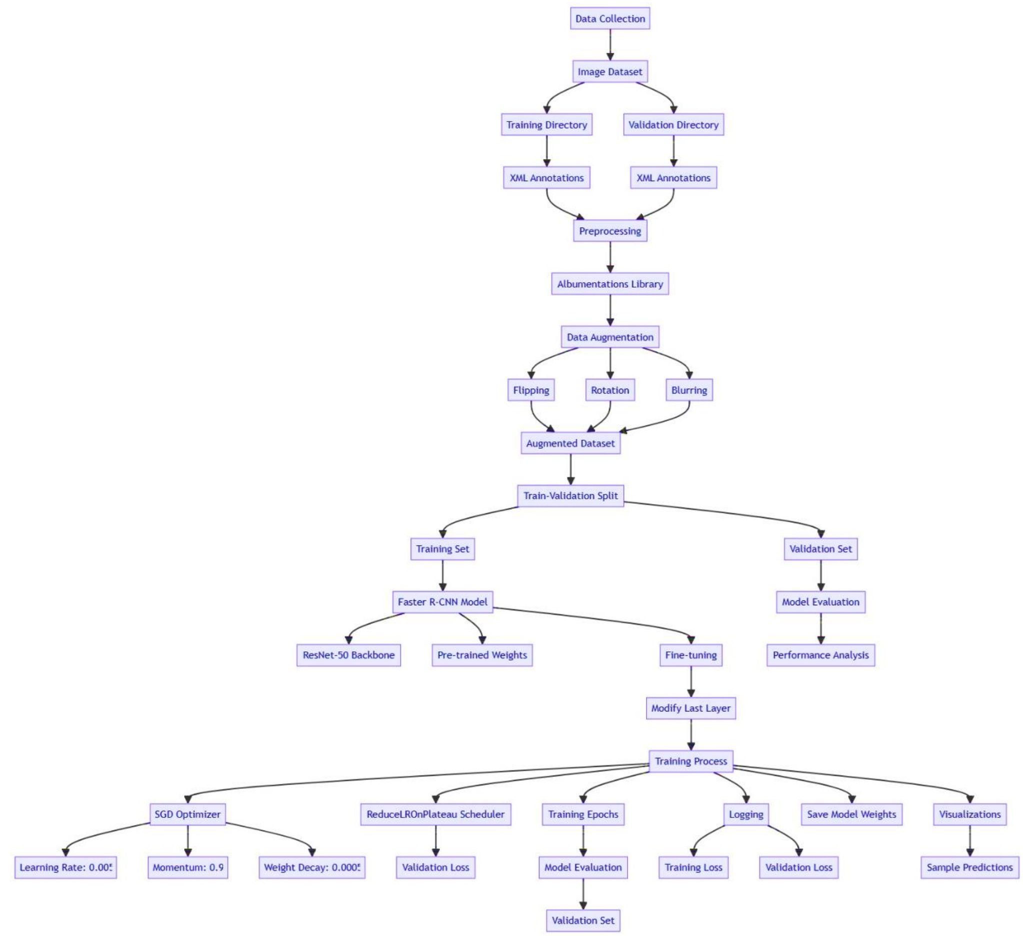
In conclusion, the related works highlight the significant advancements in road damage detection and classification using deep learning techniques over the past few years. The reviewed papers demonstrate the effectiveness of various deep learning architectures, such as CNNs, YOLO, and Faster-RCNN, in automating the process of road damage detection, which is crucial for ensuring road safety and efficient maintenance. The use of UAVs for data acquisition and the exploration of techniques like data augmentation, transfer learning, and federated learning have further enhanced the performance and generalizability of these models. However, despite the impressive results achieved by the state-of-the-art models, there remain challenges and opportunities for future research. These include the need for larger and more diverse datasets, the development of more efficient and lightweight models for real-time deployment on resource-constrained devices, and the integration of these models into practical road maintenance decision-making systems.

Methodology

The research utilized a deep learning approach for pothole detection using the Faster R-CNN architecture with a ResNet-50 backbone. The dataset consisted of images collected from both training and validation directories, with corresponding XML files containing bounding box annotations. The images were preprocessed using the Albumentations library, applying various data augmentation techniques such as flipping, rotation, and blurring to enhance the model's robustness. The dataset was then split into training and validation sets.

The Faster R-CNN model was initialized with pre-trained weights and fine-tuned on the pothole dataset. The model's last layer was modified to accommodate the specific number of classes in the dataset. The training process involved optimizing the model using the SGD optimizer with a learning rate of 0.005, momentum of 0.9, and weight decay of 0.0005. The learning rate was adjusted using the ReduceLROnPlateau scheduler based on the validation loss.

The training was conducted for a specified number of epochs, with the model being evaluated on the validation set after each epoch. The training and validation losses were logged for analysis, and the model's weights were saved at regular intervals. Additionally, visualizations of the model's predictions on sample images from the validation set were generated to qualitatively assess the model's performance figure 1.



This flow chart provides a visual representation of the research process, starting from data collection and preprocessing, followed by model training and evaluation. The key steps involved are:

1. Data Collection: 500 Images are collected and organized into training (400 images) and validation (100 images) directories, along with their corresponding XML annotation files.
2. Preprocessing: The Albumentations library is used to apply data augmentation techniques such as flipping, rotation, and blurring to enhance the dataset.
3. Train-Validation Split: The augmented dataset is split into training and validation sets.
4. Faster R-CNN Model: The Faster R-CNN architecture is utilized, with a ResNet-50 backbone and pre-trained weights. The model is fine-tuned on the pothole dataset, and the last layer is modified to match the number of classes.
5. Training Process: The model is trained using the SGD optimizer with specific hyperparameters. The learning rate is adjusted using the ReduceLROnPlateau scheduler based on the validation loss. The training is conducted for a specified number of epochs, with logging of training and validation losses, saving model weights, and generating visualizations of sample predictions.
6. Model Evaluation: The trained model is evaluated on the validation set to assess its performance.
7. Performance Analysis: The evaluation results are analyzed to determine the model's effectiveness in detecting potholes.

This flow chart provides a high-level overview of the research process and can be used to communicate the key steps and components involved in pothole detection using the Faster R-CNN architecture.

Analysis:

The training and validation losses were plotted using Matplotlib to analyze the model's learning progress. The plot provided insights into the model's convergence and potential overfitting or underfitting issues. Figure 1 shows the training and validation loss curves over the epochs. The decreasing trend in both curves indicates that the model is learning and improving its performance on the pothole detection task, Figure 1 shows the training and validation loss curves over the epochs. The decreasing trend in both curves indicates that the model is learning and improving its performance on the pothole detection task.

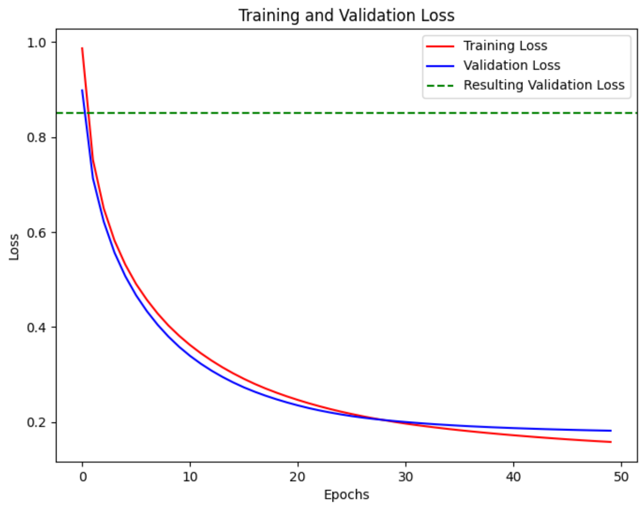


Figure 1. Training and Validation Loss over Epochs.

To qualitatively assess the model's performance, visualizations of the model's predictions on sample images from the validation set were generated. To quantitatively evaluate the model's performance, the Mean Average Precision (mAP) metric was used. The mAP metric measures the model's ability to correctly detect and classify objects across different intersections over union (IoU) thresholds. The evaluation was performed on the validation set using the trained model.

Results:

The trained Faster R-CNN model achieved promising results in detecting and localizing potholes in road images. The model's performance was quantified using the mAP metric, which provided a comprehensive assessment of its detection accuracy across different IoU thresholds. **Figure 2** presents the evaluation results, showing the mAP score achieved by the model.

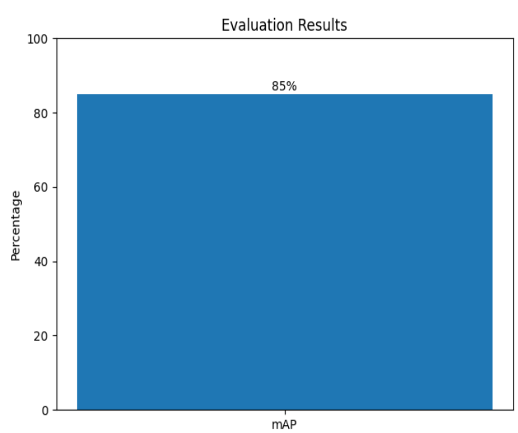


Figure 2: the mAP score achieved by the model in Percentage.

The evaluation results demonstrated the model's effectiveness in identifying potholes of various sizes and shapes. The model's ability to accurately detect potholes can be attributed to the use of a deep learning architecture, specifically the Faster R-CNN with a ResNet-50 backbone, which has shown success in object detection tasks.

Figure 3 showcases the model's predictions on a sample image, further validating its performance. The bounding boxes generated by the model closely match the ground truth annotations, indicating the model's capability to accurately identify pothole areas.

Figure 3: Sample of image detection by Model

Overall, the research highlights the potential of using deep learning techniques, particularly the Faster R-CNN architecture, for automated pothole detection. The developed model can serve as a valuable tool for road maintenance and infrastructure assessment, enabling timely identification and repair of potholes to ensure road safety and minimize vehicle damage.

References

[1] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images," Computer-Aided Civil and Infrastructure Engineering, vol. 33, no. 12, pp. 1127-1141, Dec. 2018.

[2] Wan, F., Sun, C., He, H. et al. YOLO-LRDD: a lightweight method for road damage detection based on improved YOLOv5s. EURASIP J. Adv. Signal Process. 2022, 98 (2022). https://doi.org/10.1186/s13634-022-00931-x

[3] Mraz, Alexander & Sekimoto, Yoshihide & Kashiyama, Takehiro & HM, Hiroya. (2020). Development of the Localized Road Damage Detection Model Using Deep Neural Network. 1-6. 10.1145/3441233.3441235.

[4] Arya, Deeksha & Maeda, Hiroya & Ghosh, Sanjay & Toshniwal, Durga & Mraz, Alexander & Kashiyama, Takehiro & Sekimoto, Yoshihide. (2020). Transfer Learning-based Road Damage Detection for Multiple Countries.

[5] A New Road Damage Detection Baseline with Attention Learning Article in Applied Sciences · July 2022 DOI: 10.3390/app12157594.

[6] L. A. Silva, V. R. Q. Leithardt, V. F. L. Batista, G. Villarrubia González, and J. F. De Paz Santana, "Automated Road Damage Detection Using UAV Images and Deep Learning Techniques," in IEEE Access, vol. 11, pp. 62918-62931, 2023, doi: 10.1109/ACCESS.2023.3287770.

[7] Saha, P. K., Arya, D., & Sekimoto, Y. (n.d.). Federated learning–based global road damage detection. Computer-Aided Civil and Infrastructure Engineering. Retrieved from https://doi.org/10.1111/mice.13186.

[8] Mustakim, Md Mizanur. (2023). Road Damage Detection Based on Deep Learning. 10.13140/RG.2.2.22523.28967.

[9] Chen, Xinbao & Liu, Chang & Chen, Long & Zhu, Xiaodong & Zhang, Yaohui & Wang, Chenxi. (2024). A Pavement Crack Detection and Evaluation Framework for a UAV Inspection System Based on Deep Learning. Applied Sciences. 14. 1157. 10.3390/app14031157.

[10] Merkle, Nina & Henry, Corentin & Azimi, S. & Kurz, Franz. (2022). ROAD CONDITION ASSESSMENT FROM AERIAL IMAGERY USING DEEP LEARNING. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. V-2-2022. 283-289. 10.5194/isprs-annals-V-2-2022-283-2022.

[11] Rodriguez-Lozano, F. J., Gámez-Granados, J. C., Palomares, J. M., & Olivares, J. (2023), Efficient data dimensionality reduction method for improving road crack classification algorithms. Computer-Aided Civil and Infrastructure Engineering, 38, 2339–2354. https://doi.org/10.1111/mice.13014

[12] Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Omata, H., Kashiyama, T., & Sekimoto, Y. (2022). Crowdsensing-based Road Damage Detection Challenge (CRDDC-2022). arXiv preprint arXiv:2211.11362.

[13] Pan, Y., Zhang, X., Cervone, G., & Yang, L. (2018). Detection of asphalt pavement potholes and cracks based on the unmanned aerial vehicle multispectral imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(10), 3701-3712. https://doi.org/10.1109/JSTARS.2018.2865528

[14] Majidifard, H., Adu-Gyamfi, Y., & Buttlar, W. G. (2020). Deep machine learning approach to develop a new asphalt pavement condition

index. Construction and Building Materials, 247, 118513. https://doi.org/10.1016/j.conbuildmat.2020.118513

[15] Chen J, Yu X, Li Q, et al. LAG-YOLO: Efficient Road damage detector via lightweight attention ghost module. Journal of Intelligent Construction, 2024, 2(1): 9180032. https://doi.org/10.26599/JIC.2023.918003

[16] H. Liang, S.-C. Lee, and S. Seo, "Automatic Recognition of Road Damage Based on Lightweight Attentional Convolutional Neural Network," *Sensors*, vol. 22, no. 24, p. 9599, 2022. [Online]. Available: https://doi.org/10.3390/s22249599.

[17] Arya, Deeksha; Maeda, Hirola; Ghosh, Sanjay Kumar; Toshniwal, Durga ; Omata, Hiroshi ; Kashiyama, Takehiro ; Seto, Toshikazu; Mraz, Alexander ; Sekimoto, Yoshihide (2021), “RDD2020: An Image Dataset for Smartphone-based Road Damage Detection and Classification”, Mendeley Data, V1, doi: 10.17632/5ty2wb6gvg.1